

```
(5')
       SDLEQERRAKEKLQEQQ
    1
       SDLEQDRLAKEKLQEQQ
    35 SDLEQERLAKEKLQEQQ
    52
        SDLEQERRAKEKLQEQQ
    69
        SDLEQERRAKEKLQEQQ
        SDLEQDRLAKEKLQEQQ
    103 SDLEQERRAKEKLQEQQ
    120 SDLEQER KAKEKLQEQQ
    137 SDLEQERLAKEKLQEQQ
    154 SDLEQERRAKEKLQEQQ
    171 SDLEQERRAKEKLQEQQ
    188 SDLEQERRAKEKLQEQQ
    205 RDLEQ
    210 RKADTKKNLERKKEHGDILAEDLYGRLEIP
    240 AIELPSENERGYYIPHQSSLPQDNRGNSRD
    270 SKEISIIEKTNRESITTNVEGRRDIHKGHL
                            316 (3') (SEQ ID NO:31)
    300 EEKKDGSIKPEQKEDKS
```

AAAGCGATCTAGAACAAGAGAGGCGTGCTAAAGAAAGTTGCAAGAACAAC **AAAGCGATTTAGAACAAGAGACTTGCTAAAGAAAGTTGCAAGAACAAC** AAAGCGATCTAGAACAAGAGAGGCGTGCTAAAGAAAGTTGCAAGAACAAC AAAGCGATTTAGAACAAGAGAGGCGTGCTAAAGAAAGTTGCAAGAACAAC AAAGCGATTTAGAACAAGAGACTTGCTAAAGAAAGTTGCAAGAACAAC <u> AAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC</u> **AAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC** AAAGCGATTTAGAACAAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAC AAAGCGATTTAGAACAAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAC AAAGCGATTTAGAACAAGAGAGGTGCTAAAGAAAAGTTGCAAGAACAAC AAAGCGATTTAGAACAAGAGAGGCGTGCTAAAGAAAGTTGCAAGAACAAC **AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGC** AAAGAGATTTAGAACAA 256 307 (5') 1 154 358 205 409 460 511 562 613

FIGURE 2

CAGGACAACAGAGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAA AAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAGGGATATA CATAAAGGACATCTTGAAGAAAGAAAGATGGTTCAATAAAACCAGAACAA

950 (3') (SEQ ID NO: 32)

AAAGAAGATAAATCT

834

ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTT CCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT

AGGAAGGCTGATACGAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGAT

630

732

681

E

RDELFNELLNSVDVNGEVKENILEESQVNADIFNSLVKSVQQEQQ
HNVEEKVEESVEENDEESVEENVEENVEENDDGSVASSVEESI
ASSVDESIDSSIEENVAPTVEEIVAPTVEEIVAPSVVEKCAPSVE
ESVAPSVEESVAEMLKER (SEQ ID NO. 24)

NSRDSKEISIIEKTNRESITTNVEGRRDIHK

DELFNELLNSVDVNGEVKENILEESQ LEESQVNDDIFSNSLVKSVQQEQQHNV VEKCAPSVEESVAPSVEESVAEMLKER **FIGURE 5**

LSA-TER (SOO ID NO: 23)

7295-NRI (SEC. 10 NO. 26) 7295-NRII (SEC. 10 NO. 27) 7295-RED (SEC. 10 NO. 28)

NUCLEOTIDE SEQUENCE OF THE LSA GENE 5' END

(NON-CODING 5' END)

1 AAAGTATACATCTTCCTTCTTTACTTCTTAAA

(CODING 5' END, UNIQUE)

- 33 ATGAAACATATTTTGTACATATCATTTTACTTTATCCTTGTTAATTTATTG
- 84 ATATTTCATATAAATGGAAAGATAATAAAGAATTCTGAAAAAGATGAAATCA
- 135 TAAAATCTAACTTGAGAAGTGGTTCTTCAAATTCTAGGAATCGAATAAATGA
- 186 GGAAAATCACGAGAAGAAACACGTTTTATCTCATAATTCATATGAGAAAACT
- 237 AAAAATAATGAAAATAATAAATTTTTCGATAAGGATAAAGAGTTAACGATGT
- 288 CTAATGTAAAAAATGTGTCACAAACAAATTTCAAAAGTCTTTTAAGAAATCT

- '441 AAGGGCAAGACGAAAACAGACAAGAAGATCTTGAAGAAAAAGCA

(CODING 5' END, repetitive)

- 492 GCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGACGT
- 543 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGAC<u>T</u>T
- 594 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT
- 645 GCTAAAGAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGAC<u>T</u>T
- 696 GCTAAAGAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT
- 747 GCTAAAGAAAAGTTGCAAGAACAACAAGCGATTTAGAACAAGAGAGACGT
- 798 GCTAAAGAAAGTTGCAAGAACAACAAGCGATTTAGAACAAGAGAGAC<u>T</u>T 849 GCTAAAGAAAAGTT<u>A</u>CAAGAGCAGCAAAGCGATTTAGAACAAGATAGACTT
- 900 GCTAAAGAAAGTTGCAAGAACAACAAGCGATTTAGAACAAGAGAGACGT 951 GCTAAAGAAAGGTTGCAAGAACAACAAGCGATTTAGA 988 (SEQ ID NO: 34)

LSA.5'/ATG - -> 1-phase Translation

956 b.p.

DNA sequence

linear (564 10 MO: 36) AAGCGATTTAGA

GAA his arg AAG lys CTT AAA AGT lys ser AAG TTG his GAT asp leu CAC AAA TTA AAT # asn AAA TTC phe AAC lys ATA AAA TCT / ile lys ser GAG AAG / glu lys TTT ATC CTT GTT AAT TTA TTG phe ile leu val asn leu leu TTC phe GAA AAT AAA 1 glu asn lys 1 CAA ACA AAT 1 gln thr asn p AAA AAT AAT / CAC his GAG GAA AAT (glu glu asn h ATC ile glu GTG TCA val ser GAA glu ATA AAG AAT TCT GAA AAA GAT GAA ile lys asn ser glu lys asp glu 151 / 51 CTT AAA AAT (31 / 11 G TAC ATA TCA TTT TAC TTT AT u tyr ile ser phe tyr phe il a1 / 31 AAA AAT lys asn asn AAT asn AAA AAT lys 7 91 ile 71 ATA 151 / JG AAT CGA Alr asn arg il GTA AAA ACT val GAG , AAT AGG ' GAG arg TCT ATG met TAT GTT ser (SEG ID NO: 39)ATG AAA CAT ATT TTG -ATAile ACG / TCA ser AAT asn **GGT** AAG / TCA TTA ser asn AAT leu CAT AAT GGA GAG ser **GGT** AAA (SEQ UD NO 38) met 1 61 / 61 / ATA A ATA A 121 121 AGT G ser 181 TTA 1eu 241 GAT asp 301 TTA

FIGURE 7A

lys

lys

len

phe

asu

val

len

NUCLEOTIDE SEQUENCE OF THE LSA GENE 3' END

(CODING 3' END, REPETITIVE)

- 1 CAAGAACAACAAGCGATCTAGAACAAGAGAGACGT
- 37 GCTAAAGAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT
- 88 GCTAAAGAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTT
- 139 GCTAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGT
- 190 GCTAAAGAAAGTTGCAAGAACAACAAGCGATTTAGAACAAGAGAGACGT
- 241 GCTAAAGAAAGTTGCAAGAACAACAAGCGATTTAGAACAAGATAGACTT
- 292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGT
- 343 GCTAAAGAAAGTTGCAAGAACAACAAGCGATTTAGAACAAGAGAGACGT
- 394 GCTAAGAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTT
- 445 GCTAAAGAAAGTTGCAAGAACAACAAGCGATTTAGAACAAGAGAGACGT
- 496 GCTAAAGAAAAGTTGCAAGAACAACAAGCGATTTAGAACAAGAGAGACGT
- 547 GCTAAGAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT
- 598 GCTAAAGAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA

(CODING 3' END, UNIQUE)

- 640 AGGAAGGCTGATACGAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGAT
- 691 ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTT
- 742 CCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT
- 793 CAGGACAACAGAGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAA
- 844 AAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATA
- 895 CATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA
- 946 AAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAAT
- 997 ATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT
- 1048 GCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAAGATGATGAAGACT
- 1099 TAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAA
 - 1150 ACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATT
 - 1201 TAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAG
 - 1252 AAAAAATAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTA
 - 1303 AACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATG
 - 1354 ATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATA
 - 1405 TATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAG
 - 1456 ATATAACTAAATATTTTATGAAACTATAA (stop) (SEQ ID No: 39)

(NON-CODING 3' END)

1485 AAGGTTATATATTT 1498

1-phase Translation ^ LSA.3'.ALL

(SEC ID NO: 41) (SEQ ID NO: 40) CAAGAACAACAA b.p. 1496 DNA sequence

CAA asp CAA arg GAA GAT CGT GAA glu AGC GAA glu AGA arg AAA Ser CAA gln CAA gln CTA GAG glu GCT leu GAA AAG TTG glu lys leu asb CAA gln CAG gln GAT CT GAG AGA glu AGC ser GAA glu CAA CAA GAT gln TTA len gln AAA lys TTA asb S CAA gln leu **GAT** AAG lys ala GAA glu GAA AGC ser GAA glu CAA 190 TTA arg CAA gln AGA arg AAA TTG len GAT CAA gln lys GAG glu SCT lys glu AAG GAA AGC glu gln / 31 glu GAA CAA len GAA CAA CAA AGA AAA CAA arg lys leu 31 CTA len GAA asb AAG **GAT** lys 271 asp GAT CAA gln len GAA glu CAA gln CT GAA glu AGA AAA arg lys CAA AGC ser gln GAG TTA glu GCT ala AAG len gln 21 (SEQ 10 NOT 42) CAA GAA CAA (SEQ 10 NOT 43) gln glu gln 61 / 21 CAA GAT asp gln GAA CGT arg GAA AGA AAA Ser CAA leu 181 GAG 8lu 241 GCT gln 121 TTA

CAA

TTG leu

AAG lys

GAA

AAA

CCT CT

CGT

AGA

GAG

CAA

GAA

SA gln

CAG

GAG

glu

leu

lys

301

glu

lys

ala

arg

nlg

leu

lys

leu

arg

asp

gln

glu

len

asp

ser

GAA AAA glu lys CAG ATC GTG GAT gln ile val asp CCA AAT GAT AAA AGT pro asn asp lys ser GAA AAG (glu lys § phe asp gly asp asn glu ile leu gln ile 491 TAT TTT ATG AAA CTA TAA AAG GTT ATA TAT tyr phe met lys leu<mark>[OCH lys vat The tyr</mark> the lys aspasnasnashe lys proasnas 451 TAAA AAT GAT AAG CAG GTT AAT AAG GA rlys asnasplys glnvalasnlys gl 471 A TTT GAC GGA GAC AAT GAA ATT TTA CA e phe asp gly aspasngluile leugl AAA TTT AAA ACA AAG GAT AAT AAT lys thr lys asp asn asn 1351 / 451 AAA 1 lys 1 1411 lys phe TCT GAA GAT asp tyr len 421 AAG AAA 1 lys lys 1 441 GAG CAT A glu his i 461 lys ser [.] 481 glu AAA TCA tyr asp 1381 / TTC ATA / phe ile l 1261 / AAA GGA / lys gly / 1321 / TAT GAT (phe ile 1441 / GAG TTA

FIGURE 9D

* Strike from

1-phase Translation ^ LSN.3'ST0P

(SEG 10 MO 145) ATGAAACTATAA (SEG ID NO: 44) CAAGAACAACAA b.p. 1482 DNA sequence

GAA glu CAA gln TTG len GAA AAG galu lys AAA AGA CGT GCT / arg arg ala GAG glu GAA CAA glu gln 31 CTA len asb GAT CAA AGC ser gln $1\cdot /1$ (569 ID NO: 46) CAA GAA CAA (**seq.10 no:** 47)gln 61

asp CAA GAT ser CAA gln gln CAG GAG glu CAA gln TTA leu lys AAG GAA / AAA lys CCT leu 51 AGA CTT CÁA gln GAA glu TTA len asb AGC GAT ser CAA

GAA glu leu asb GAT AGC ser gln CAA CAA gln GAA glu CAA gln TTG leu lys AAG glu AAA GAA lys leu AGA arg nlg GAG CAA gln 61 **GAA** gln 121 171 181 181 6AG 6AG 8lu 241

AGC ser CAA / CAA gln **GAA** TTG CAA len AAG GAA glu AAA lys ala arg CGT AGA

AGA glu GAT leu GAT **S** GAA TTA GAT AGC CAA CAA lys 271 GAA CAA TTG AAG GAA AAA

arg

CGT

AGA arg

GAG glu

gln

CAA

GAA

TTA

GAA

GCT AAA

CTT

lys

len arg asb gln nlg len asb ser leu lys ala 301 AAG lys

S C leu AAG GAA glu AAA (ala CGT GCT arg GAG AGA glu arg CAA gln **GAA** glu leu gln SA CAG gln GAG glu

FIGURE 10A

GAA AAA glu lys GGA GAC AAT GAA ATT TTA CAG ATC GTG GAT gly asp asn glu ile leu gln ile val asp AA AAA TAT AAA AAT GAT AAG CAG GTT AAT AAG GAA AAG GAA AAS Iys lys tyr lys asn asp lys gln val asn lys glu lys glu 1411 / 471

TT CAT ATA TTT GAC GGA GAC AAT GAA ATT TTA CAG ATC GTG he his ile phe asp gly asp asn glu ile leu gln ile val 1471 / 491

TA ACT AAA TAT TTT ATG AAA CTA TAA AAG GTT ATA TAT le thr lys tyr phe met lys leu GCH lys val ile tyr. AAA lys TTT phe ATA ile 1321 / 441
TAT GAT GAG CAT ATT AV
tyr asp glu his ile ly
1381 / 461
TTC ATA AAA TCA TTG T
phe ile lys ser leu pl
1441 / 481
GAG TTA TCT GAA GAT A'
glu leu ser glu asp i

FIGURE 10D

* Delete from Figure